DOCUMENT RESUME

ED 390 940 TM 024 571

AUTHOR Harker, Richard

TITLE Analysis of School Effects on School Certificate Results through the Use of Hierarchical Linear Models.

PUB DATE Apr 95

NOTE 27p.; Paper presented at the Annual Meeting of the American Educational Research Association (San Francisco, CA, April 18-22, 1995).

PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC02 Plus Postage.

DESCRIPTORS *Academic Achievement; *Educational Quality; Effective Schools Research; English; Ethnic Groups; Foreign Countries; *Institutional Characteristics; Mathematics Education; *School Effectiveness; Science Education; *Student Characteristics

IDENTIFIERS *Hierarchical Linear Modeling; *New Zealand

ABSTRACT

School effectiveness was studied in New Zealand schools, concentrating on Type A effects (given average background characteristics, how well would a student perform in any particular school?) and Type B effects (given similar student populations, are some schools more effective in achieving specified results than others?). Hierarchical linear modeling was used on data for 37 schools and 5,391 students to establish that 18.2% of total variance in mathematics, 14.3% in science, and 14.6% in English was related to the characteristics of the schools rather than inter-individual variability within the population sampled. Some schools were significantly more effective than others in that school means for the School Certificate examinations varied significantly between schools. However, a considerable proportion (around 70%) of the between-school variance in all three subject areas was due to the initial ability and social and ethnic characteristics of the student populations. This Type A effect varied. Controlling for these factors (to produce a better Type B estimate) shrank the differences between schools and changed the rank order of effective schools from that based on raw scores alone. (Contains 6 figures, 10 tables, and 19 references.) (SLD)

Paper to:
1995 Annual Meeting of American Educational Research Association
April 18-22, San Francisco.

Analysis of School Effects on School Certificate Results Through the Use of Hierarchical Linear Models.

Richard Harker[1]
Department of Policy Studies in Education
Massey University
Palmerston North
New Zealand

1. Introduction

The question of school effectiveness is a perennial one in educational circles. It is also an important political issue in an era of accountability in which a market philosophy of 'value for money' is a crucial determinant of policy and resource allocation. However, in many respects, schools are rather similar to each other while families are rather more heterogeneous - i.e., family background variables have been found to account for a lot more variance than school based variables (Rutter, 1983). Estimates of the power of school environments to modify the effects of family background vary widely, depending on the criterion selected as the outcome variable and the way the school effect is measured. The outcome criterion is important, since for some areas of achievement (such as language), learning takes place in a wide variety of contexts outside the school environment, while in other areas (such as Mathematics and Science), learning is much more directly school based. It would not be surprising then if Mathematics and Science results were much more closely tied to school characteristics than Language tests. In general terms however, 'School Effects' research tries to tease out answers to the following questions:

* for a specified criterion variable, what proportion of total variance is systematically related to characteristics of schools, as opposed to the general variability that occurs between individuals?

---

1 I am indebted to: Roy Nash, the Director of the project from which the data used in this paper are generated; to Charles Lawoko and to J. Douglas Willms for their comments and advice at various stages of the work.

* are some schools more effective than others, or is the 'school effect' much the same in all schools?
* if some schools are more effective than others, is it to do with some characteristics of the school or more to do with the characteristics of the student population that attend the school (the 'school-mix' effect)? This can be teased out in two ways - which Willms and Raudenbush call Type A and Type B effects (1989:213);
* given average background characteristics, how well would a student perform in any particular school? (Type A effect);
* given similar student populations, are some schools more effective in achieving specified results than others? (Type B effect).

Clearly, parents choosing a school for their child, will be interested in Type A effects, while administrators and officials, looking for accountability and 'value for money', are more interested in Type B effects. One of the main objectives of the 'Progress at School' project, funded by the Ministry of Education, was to seek answers to these questions for New Zealand schools. In seeking answers to these questions we have used the technique called Hierarchical Linear Modelling, which allows the simultaneous control of individual and school level data.

Hierarchical data structures are those in which observations on individuals are nested within larger organisational units such as business firms, regions or schools. The data set reported in this study is of this kind, whereby data on 5391 students are nested within 37 schools. A large number of variables exist on the individual students, their backgrounds, present characteristics and test performances. In addition, data has been collected on the schools attended by these students which describe some of their characteristics and catchment areas. The analysis of such data presents a number of difficulties, the resolution of which has had to wait upon the development of sufficiently powerful statistical algorithms to allow the full utilisation of the data set

without reducing (or aggregating) the data at one level into the terms of the other. (See Bryk and Raudenbush[2] 1992:83 et seq. for further details)

The HLM software (Hierarchical Linear Modelling), developed by Bryk, Raudenbush and Congdon (1994), is used in the analysis of the 'School Effects' data gathered during 1991-4 as part of the 'Progress at School' research project, the preliminary analysis of which was reported in Nash and Harker (1994) and Harker and Nash (1994). A two-level model is used in the analysis reported here, and the logic and basic statistical models are described in Chapter 2 of (B&R). A more basic, non-technical introduction is provided in Paterson (1991).

2. The 'Progress at School' data

A full report on the data collected for the first phase of this project can be found in the main report (Nash and Harker, 1994). This paper uses the results of the National School Certificate Examinations, specifically the results of the English, Mathematics and Science examinations. Typically this examination is conducted at the end of Form 5 (Grade 10), but some few sit the exam (particularly Mathematics) at the end of Form 4 (Grade 9). Not all pupils sit these examinations, as some have already left school and others have not made sufficient progress in their work. In addition, some students do not sit the examination until the end of the sixth Form (grade 11), and the results of the latter group are not included in this preliminary analysis. Hence the numbers of students in our analyses vary, depending on the number in each subject for whom we have results.

The characteristics of the schools involved in the study are outlined in Table 2.1.

2 This book will be frequently cited in the report, and will be referred to as (B&R).

Table 2.1: Social characteristics and attainment scores of schools.

ID	Type	Third Form N	School Means: SES*	Percent Maori	Percent Pac.Is.	Initial Ability	S.Cert Maths	S.Cert Science	S.C Eng
01	co-ed	119	-3.55	31.1	6.7	-0.42	44.5	52.9	45.6
02	co-ed	148	-3.06	16.2	0.7	0.09	56.3	60.6	51.2
03	boys	128	-1.30	8.6	1.6	0.92	73.8	64.7	63.0
04	co-ed	203	-2.95	24.1	16.3	-0.24	51.5	52.6	51.5
05	co-ed	96	-3.76	37.5	18.8	-0.58	41.6	42.8	44.9
06	co-ed	228	-3.25	19.7	0.9	-0.08	49.5	53.2	51.7
07	co-ed	177	-2.52	19.2	3.4	0.17	53.7	56.1	51.3
08	co-ed	91	-3.78	36.3	2.2	-0.37	44.7	46.9	48.0
10	co-ed	164	-3.54	28.7	1.8	0.02	49.1	56.9	54.1
11	co-ed	163	-3.46	23.3	0.6	0.06	53.7	57.4	55.3
12	co-ed	125	-3.13	32.8	4.8	-0.27	46.3	50.3	49.7
13	co-ed	237	-2.95	24.1	6.3	-0.20	47.6	49.0	48.9
14	girls	51	[illegible].65	21.6	5.9	0.10	52.2	54.0	60.6
15	girls	169	-3.40	24.3	0.6	0.03	52.9	55.4	57.0
16	boys	186	-3.28	10.8	24.7	-0.19	45.8	46.3	41.1
17	co-ed	68	-2.69	22.1	0.0	0.33	48.8	53.7	50.9
18	co-ed	128	-3.35	25.8	0.0	0.03	59.1	58.3	53.2
19	co-ed	86	-3.57	30.2	1.2	-0.54	52.1	53.4	41.6
20	co-ed	95	-3.87	42.1	14.7	-0.36	43.1	45.2	50.7
21	co-ed	118	-3.71	21.2	0.8	-0.02	48.0	53.3	56.0
22	boys	168	-2.75	13.8	0.0	0.11	53.8	54.6	51.2
23	co-ed	271	-4.18	17.3	50.9	-0.75	36.9	42.7	41.4
24	co-ed	191	-3.13	16.2	19.9	-0.13	48.4	51.5	47.9
25	co-ed	69	-3.29	20.3	5.8	-0.21	44.0	42.8	48.0
26	girls	134	-3.07	11.2	6.0	0.34	48.4	56.6	58.7
27	co-ed	104	-3.26	32.7	1.0	0.04	43.0	49.9	50.0
28	girls	237	-2.60	14.8	1.3	0.29	60.7	61.1	59.1
29	boys	124	-3.13	9.7	0.8	0.09	53.2	51.5	49.1
30	co-ed	91	-3.31	19.8	3.3	0.10	52.3	61.7	49.0
31	boys	219	-2.63	17.4	0.9	0.32	58.9	61.0	59.5
32	co-ed	74	-3.14	35.1	4.1	-0.21	56.2	55.6	49.7
33	boys	277	-1.30	9.0	4.3	0.58	70.7	65.8	58.5
34	co-ed	87	-2.87	20.7	2.3	0.17	52.8	55.2	52.4
35	co-ed	138	-3.40	29.0	0.0	-0.03	54.7	56.8	48.1
36	girls	133	-3.14	18.0	34.1	0.05	47.1	49.2	53.1
37	girls	141	-3.40	46.1	3.5	-0.08	47.1	51.4	57.4
38	co-ed	133	-3.45	26.3	1.5	-0.18	43.6	52.0	50.2
Total means		145.7	-3.13	23.2	6.5	-0.03	50.9	53.6	51.6

* The Elley-Irving Scale has had 1 subtracted from each value, then multiplied by -1. The revised scale now runs from -5 (Lowest SES category [E-I. 6]) to 0 (Highest SES category [E-I.1]). This procedure gives the intercept point (0) a meaning, and produces positive correlations, in line with conventional practice.

BEST COPY AVAILABLE

3. The HLM Analysis

In the analyses which follow, a variety of models are tested, and in each case are run against the scores attained by students in the Maths, Science and English examinations separately. As a first step, a fully unconditional model was tested[3], as it produces a number of useful outputs:

1. a point estimate and the 95% confidence interval for the grand mean of the dependant variable;
2. the outcome variability at each of the two levels of data - i.e., within-group variability and between-group variability;
3. the intraclass correlation coefficient which measures the proportion of the variance in the outcome that is between the level-2 units (schools);
4. a reliability estimate of the sample means as estimates of each school's true mean (averaged across all 37 schools);
5. a test of the null hypothesis that all schools have the same achievement mean, i.e., that level-2 variance is not significantly greater that zero. This is evaluated with the chi squared test (see(B&R):63-64).

For the three outcome variables in our data set, a summary of the results of running the fully unconditional model are reported in Table 3.1, which provides answers to the questions above. The most important analytic point to note from Table 3.1 is that in a fully unconditional model, i.e., without taking any possible intervening variables into account, differences between schools are significantly different from zero, and account for 18.25 percent of the total variance in Maths[4], 14.34 percent in Science scores, and 14.57 percent of the English scores. While the great majority of the variability is between individual pupils, rather than between schools, the difference between schools

---

3 Under these conditions the linear model reduces to (B&R:17; 33-35):
$Y_{ij} = \beta_{0j} + r_{ij}$ (for level-1); and $\beta_{0j} = \Gamma_{00} + u_{0j}$ (for level-2).

4 This compares with a level of 19 % reported for a sample of 160 schools in the United States of America, with Math achievement as the dependent variable (B&R:104), and within the range reported for their U.K. study by Smith and Tomlinson (1989)

Table 3.1: Fully unconditional model results.

Outcome	Maths	Science	English
Grand mean	50.89	53.60	51.62
Std. Error	1.22	0.95	0.87
Level-2 variance	53.04	31.69	27.06
Level-1 variance	237.70	189.27	158.63
Intraclass correl.[a]	0.1824	0.1434	0.1457
Reliability est.	0.965	0.954	0.955
chi-squared	1453.85	1022.25	1012.15
probability	<0.000	<0.000	<0.000

[a] Divide level-2 variance by the sum of level-2 and level-1 variance.

is highly significant and of a similar order of magnitude in all three achievement areas.

The next step is to analyse the between individual variance component. To do this, the three criterion variables need to be regressed against level-1 variables. This allows differences between schools to be adjusted for differences at the individual pupil level, since we already know that schools vary greatly in terms of the ability level of their pupil intakes (see the summary data in Table 2.1).

3.1. Type A effects

Instead of dumping all possible level-1 predictors into the model and working backwards by progressively eliminating non-significant effects from the model, (B&R) recommend a 'step-up' procedure guided by theory and by previous uni-variate analyses. It is clear from our preliminary report and from work in other countries ((B&R); Willms 1992) that the best individual predictor of achievement scores is the initial ability of the student. This is confirmed with our data set by running a regular stepwise regression on the data from which the HLM data set was generated. Table 3.2 shows the summary results of regressing likely variables against the three criterion variables.

Table 3.2: Multiple (stepwise) regression: level-1 variables

Variables	Maths		Science		English	
STEP 1						
R	zscor91	.610	zscor91	.604	zscor91	.616
R^2		*.372*		*.365*		*.379*
STEP 2						
R	ses	.629	ses	.616	sex	.646
R^2		*.396*		*.379*		*.418*
STEP 3						
R	sex	.639	maori	.624	ses	.657
R^2		*.409*		*.389*		*.432*
STEP 4						
R	otheth	.647	otheth	.630	maori	.660
R^2		*.419*		*.397*		*.435*
STEP 5						
R	maori	.652	pacisl	.633	otheth	.661
R^2		*.425*		*.401*		*.437*

All level-1 variables were entered into the regression run, and in each case, five variables made a significant contribution to variance accounted for, though a different set of five for each of the criterion variables. However, it is also clear from the data that initial ability (zscor91) has an overwhelming effect, and that despite sometimes quite high zero-order correlations between some of the other predictor variables and the three criterion variables, communality with initial ability is high.[5] These regressions confirm that initial ability is the most powerful predictor of School Certificate results, and that the addition of other variables (even though statistically significant) does not substantially improve the predictability of individual level scores, though the effects in a two level hierarchical model may be quite different. Hence the first step is to run an HLM model[6] with each of the output variables in turn (Maths,

5 'The *communality* of a variable is the variance it shares with other variables. It is the common factor variance of a ... variable.' (Kerlinger and Pedhazur 1973:362) A variable may uniquely account for a segment of the variance on some criterion which is the complement to the *communality*. However, it is usual for variables to overlap somewhat and account for the same segment of variance as some other variable. This "overlap" is the communality.

6 The model is: $Y_{ij} = \beta_{0j} + \beta_{kj}*(X_k) + r_{ij}$ [k = 1-5] (for level-1); and (for level-2)
$\beta_{0j} = \Gamma_{00} + u_{0j}$ and
$\beta_{kj} = \Gamma_{k0} + u_{kj}$

Science and English) regressed against the variables specified in Table 3.2., with all five varia entered in the level-1 equation (each centred on the grand mean), and each specified as random in the level-2 model (see(B&R):105).

As already noted, the communality of other level-1 variables with initial ability is very great, and running a 2-level model with all five variables indicated by the initial regression (Table 3.2) only minimally reduces the amount of level-1 variance accounted for by initial ability alone (from 200.2 to 196.2 for Maths; from 149.4 to 145.4 for Science; and from 119.5 to 120.2 [an *increase*] for English). In order to achieve these refinements, too large a price is paid, in that an unacceptably large number of schools are excluded due to excessive homogeneity on some of the variables involved, which is related to very small numbers of students in some categories (eg. OTHERETH). The variable SEX, while obviously important in Maths and English, cannot be included in a 2 level model due to the fact that there are 12 single sex schools in our sample of 37, hence we would loose all 12 due to a lack of variance on the variable SEX amongst the students. SES is retained even though in all three cases it drops out of contention as a contributor of unique variance once other variables (particularly the ethnic ones) are added. However, it does have consequences for Type A effects in that we are looking at performance levels for 'average' pupils. The results of revised runs which have been modelled to retain all 37 schools, are reported in Table 3.3.

Each of the coefficients reported in the 'Fixed effect' sections of Table 3.4 is *net of the others*. Thus (for example) the average school achievement in School Certificate Mathematics is shown in panel 1 as 50.8 (as in Table 3.1). The ZSCOR91 coefficient of 6.7 indicates that in an average school, and for pupils of the same ethnic and SES background, each unit increase in initial ability lead to a highly significant increase of 6.6 marks. Similarly, for pupils of the same ability and SES level, being Maori leads to a decline of 1.8 marks - a change that is also significant. The coefficients for Science

Table 3.3: Level-1 variables. Average ability.

Mathematics

Fixed Effect	Coefficient	Standard Error	T-ratio	P-value
For INTRCPT1, B0				
INTRCPT2, G00	50.803546	0.892500	56.923	[illegible].000
For ZSCOR91 slope, B1				
INTRCPT2, G10	6.578667	0.359577	18.57[illegible]	0.000
For SES slope, B2				
INTRCPT2, G20	0.094361	0.162563	0.580	0.333
For MAORI slope, B3				
INTRCPT2, G30	-1.846344	0.511961	-3.606	0.001

Random Effect		Standard Deviation	Variance Component	df	Chi-square	P-value
INTRCPT1,	U0	5.26759	27.74748	36	520.21616	0.000
ZSCOR91 slope,	U1	1.63525	2.67404	36	91.38856	0.000
SES slope,	U2	0.66159	0.43770	36	66.18462	0.002
MAORI slope,	U3	0.61674	0.38036	36	25.16050	>.500
level-1,	R	14.08777	198.46521			

Science

Fixed Effect	Coefficient	Standard Error	T-ratio	P-value
For INTRCPT1, B0				
INTRCPT2, G00	53.715195	0.682557	78.697	0.000
For ZSCOR91 slope, B1				
INTRCPT2, G10	6.901343	0.341038	20.233	0.000
For SES slope, B2				
INTRCPT2, G20	0.034402	0.152871	0.225	0.386
For MAORI slope, B3				
INTRCPT2, G30	-1.584049	0.478436	-3.311	0.003

Random Effect		Standard Deviation	Variance Component	df	Chi-square	P-value
INTRCPT1,	U0	3.99014	15.92121	36	449.00243	0.000
ZSCOR91 slope,	U1	1.65113	2.72624	36	106.50269	0.000
SES slope,	U2	0.66855	0.44695	36	76.09587	0.000
MAORI slope,	U3	1.18976	1.41554	36	38.18952	0.370
level-1,	R	12.15395	147.71860			

(cont.)

(Table 3.3 cont.)

English

Fixed Effect	Coefficient	Standard Error	T-ratio	P-value
For INTRCPT1, B0				
INTRCPT2, G00	51.691642	0.597977	86.444	0.000
For ZSCOR91 slope, B1				
INTRCPT2, G10	6.852989	0.422153	16.233	0.000
For SES slope, B2				
INTRCPT2, G20	0.038917	0.164868	0.236	0.385
For MAORI slope, B2				
INTRCPT2, G20	-[illegible] 199878	0.416434	-2.881	0.009

Random Effect		Standard Deviation	Variance Component	df	Chi-square	P-value
INTRCPT1,	U0	3.48971	12.17806	36	472.28756	0.000
ZSCOR91 slope,	U1	2.29523	5.26808	36	191.47580	0.000
SES slope,	U2	0.81388	0.66240	36	122.85201	0.000
MAORI slope,	U2	0.91498	0.83720	36	23.88371	>.500
level-1,	R	10.82995	117.28783			

and for English are to be interpreted in the same way, and show very similar patterns to that of Mathematics. The most interesting feature of these data is the way that SES ceases to be an important factor at level-1 once ability and ethnicity are controlled for.

The very high probability levels of the chi-square tests for the intercept and for initial ability indicate that highly significant differences exist between the school means, and that there are highly significant differences between the schools, particularly in the relationship of initial ability and achievement score. The ethnic variable, while a significant factor in explaining individual level achievement, does not systematically vary between schools, as is shown by the variance estimate for MAORI failing to reach a level of statistical significance for any of the achievement criteria, which is hardly surprising given the low reliability of the estimates.[7] SES shows an opposite pattern,

7 Table 3.FN6: Reliability of estimates reported in Table 3.4.

Estimates	Maths	Science	English
Ach. means (β_0)	0.933	0.915	0.912
ZSCOR91 (β_1)	0.550	0.622	0.793
SES (β_2)	0.432	0.506	0.648
MAORI (β_3)	0.038	0.164	0.128

failing to make a contribution at the individual level, but showing a highly significant amount of variability between schools.

The estimated variances of the random effects at levels 1 and 2 are shown in the second panel of each section of Table 3.3. In each case the level-1 variance has been reduced from the level reported in the random effects ANOVA model (reported in Table 3.1). These reductions are summarized in Table 3.4. Controlling for these level-1 variables has the effect of reducing the apparent variance between schools by 48 percent in Maths, 50 percent in Science and 55 percent in English. This emphasises the point made by many researchers, that comparing the differences in achievement between schools without taking into account the ability and social mix of the pupil populations results in badly biased estimates of their achievements. The actual school-by-school residuals that result from this model, which can be interpreted as Type A effects, are shown in Table 3.5. The data in the Table are shown at three points on the initial ability measure, and shows how Type A effects can vary significantly, not only

Table 3.4: Accounting for variability: Average Ability.

Outcome	Maths	Science	English
Level-1 variance*	237.70	189.27	158.63
After control for: *ZSCOR91 + SES + MAORI*			
	198.47	147.72	117.79
% accounted for+	16.50	21.95	26.10
Level-2 variance*	53.04	31.69	27.06
After control for: *differences between schools*			
	27.75	15.92	12.18
% accounted for+	47.68	49.76	54.99

* From Table 3.1
+ (B&R), p.70. (Total level-1 variance - controlled variance)/total level-1 variance.

between schools, but also in the same school depending on the ability level of the pupil. In order to estimate a Type A effect for the different ability levels of pupils, the

the initial ability variable as centred one standard deviation below the mean, at the mean, and one standard deviation above the mean, in three separate runs. All other

Table 3.5: Type A Effects: Empirical Bayes Estimates, by Ability Levels - expressed as effect sizes (residuals as proportion of s.d).

	Mathematics			Science			English		
School	Low	Average	High	Low	Average	High	Low	Average	High
01	-.11	-.16	-.21	.11	.10	.10	-.12	-.20	-.24
02	.21	.21	.20	.36	.32	.29	.05	-.06	-.16
03	.35	.54	.75	.16	.19	.24	.08	.17	.32
04	.06	.09	.00	.06	.02	-.04	-.46	.08	.03
05	-.20	-.24	-.30	-.35	-.30	-.24	-.16	-.12	-.07
06	-.01	-.04	-.03	.02	-.00	-.02	.11	.02	-.03
07	.06	.07	.07	-.02	.05	.14	-.31	-.12	.08
08	-.12	-.13	-.14	-.23	-.18	-.13	-.00	-.06	-.09
10	-.07	-.08	-.10	.25	.16	.08	.19	.17	.12
11	-.07	-.07	-.02	.18	.17	.16	.34	.20	.09
12	-.13	-.10	-.06	-.09	-.06	-.03	.07	-.01	-.05
13	-.11	-.09	-.09	-.16	-.17	-.18	-.09	-.09	-.13
14	-.04	.03	.11	-.09	-.03	.05	.25	.41	.61
15	.08	.08	.11	.06	.08	.10	.35	.30	.33
16	-.13	-.18	-.22	-.27	-.33	-.40	-.51	-.57	-.58
17	-.23	-.19	-.15	-.12	-.11	-.10	-.08	-.16	-.18
18	.42	.36	.30	.30	.23	.16	.16	.08	.03
19	.23	.20	.18	.21	.12	.03	-.49	-.28	-.12
20	-.20	-.20	-.19	-.30	-.26	-.23	.11	13	.18
21	-.07	-.14	-.16	-.06	-.02	.03	.37	.23	.19
22	.09	.09	.09	-.04	-.01	.03	.01	-.08	-.15
23	-.37	-.45	-.67	-.29	-.41	-.67	-.26	-.24	-.58
24	-.06	-.07	-.07	-.21	-.06	.09	-.15	-.18	-.18
25	-.21	-.22	-.24	-.47	-.45	-.44	-.05	-.14	-.23
26	-.18	-.21	-.22	.06	.03	.01	.06	.21	.43
27	-.37	-.36	-.31	-.30	-.20	-.08	-.14	-.12	-.07
28	.40	.35	.32	.26	.27	.30	.31	.31	.36
29	.08	.07	.06	-.02	-.13	-.25	.00	-.18	-.38
30	.02	.05	.05	.45	.37	.29	-.22	-.18	-.18
31	.17	.23	.31	.40	.30	.20	.28	.31	.39
32	.31	.28	.26	.20	.16	.11	.01	-.03	-.05
33	.51	.60	.75	.15	.31	.49	-.22	.00	.27
34	.07	.04	.00	.06	.03	-.01	.08	-.01	-.09
35	.22	.18	.18	.22	.17	.12	-.14	-.20	-.20
36	-.23	-.18	-.11	-.37	-.26	-.13	-.03	.06	.18
37	-.07	-.13	-.23	-.08	-.07	-.06	.48	.38	.23
38	-.28	-.25	-.22	-.04	-.02	.01	.03	-.02	-.07

aspects of the model were identical, except that in the case of high ability pupils the model had to exclude SES in order to retain all schools in the model. The intercept coefficients then show the expected score for pupils at that point in the scale. A problem in analysing school effects is that the scale of the effect depends on the nature of the criterion variable, its metric and distribution. In the present study, the three criteria have similar means, but rather different standard deviations.[8] Mathematics has a much broader spread than English, hence the residuals need to be adjusted to take account of the differences. This is done by expressing the residual as a fraction of the standard deviation of the outcome measure, to produce what has come to be called in the literature, an 'effect size' (Willms, 1992:43) for each school. The resulting effect sizes are shown in Table 3.5. An effect size of 0.1 in Mathematics is equivalent to 2.15 percentage points on the School Certificate scale, in Science such an effect size represents 1.85 percentage points, and in English 1.65 percentage points.

While most schools cluster around the mean, there are a number of exceptions, with effect sizes ranging from -0.67 (school 23, high ability pupils in Maths) up to 0.75 (schools 3 and 33, high ability Maths). These very high positive effect sizes (which translate into about 16 percentage points) are both associated with single-sex boys schools which clearly make a feature of Mathematics teaching since the effects in the other two subject areas, while above average, are not exceptional. A great deal of further detailed analysis could be undertaken, such as identifying schools that seem to be doing excellent remedial work, and other schools that seem to be concentrating on high ability pupils. However, caution is demanded, and the temptation to rank schools on this basis must not be given into, since for the majority of schools, clustered close to the mean, quite small changes to the annual intake, including difference due to measurement or sampling error (Willms 1992:43) could dramatically change the school's rank in the effect size distribution. However, when parents come to choose a school for their children, it is the schools with larger positive effects, that would be of

8 For Mathematics the mean is 52.6, sd =21.7; Science 54.6, sd 18.6; English 52.2, sd 16.0.

interest, and those with larger negative effects that are to be avoided, always provided these effects were shown to be stable over a number of years. Type A effects, then, in New Zealand schools are demonstrable, and in some cases, quite substantial.

3.2: Analyzing Between-school (Level-2) variance

The HLM software allows the analyst to test for the likely impact of controlling for contextual level-2 variables by running an exploratory analysis at the end of each run. For each dependent variable, the 't-to enter' estimates for contextual level-2 variables were requested with the Type A analyses reported above, and are summarised in Table 3.6. It is clear that some of them could account for a substantial part of the between school variance. School based variables such as size, tracking policy etc., are excluded in order to estimate how much of the variance is taken up by the characteristics of the pupil population, which, by subtraction, will provide an estimate of the amount of between-school variance that can be attributed to schools as such, rather than to the students that attend them. However, these t-to-enter statistics take no account of communalities, and, as (B&R) point out, are:

> only approximate because the model is doubly multivariate with errors correlated across equations and multiple predictors for each equation. They will usually provide a good indication of the next *single* variable to enter *one* of the Level-2 equations. If several variables are entered simultaneously ... the actual results may not follow the pattern suggested by these statistics. (214-5)

Table 3.6: T-to enter statistics for level-2 vars; to model the level-1 intercept.

Level-2 variable	Maths t value	Science t value	English t value
AGGZ91	4.350	3.703	2.600
SESAGG	5.758	2.730	1.228
EUROPC	3.974	5.013	1.516
MAORIPC	-2.123	-1.169	0.664
PACISPC	-2.998	-4.163	-2.086

For Mathematics the school aggregated SES mean is the strongest candidate, for Science EUROPC, and for English initial ability (AGGZ91). A first step is to run the

model with the strongest variable only on the undestanding that the remaining variables will rearrange themselves on the basis of their communalities with that strongest variable and the variables included in level-1. In this way all possible level-2 variables (excluding all variables which relate to school policy and practice) can be examined for a unique contribution, to arrive at the most parsimonious model that maximises the level-2 variance accounted for. However, in estimating Type B effects, all of the contextual variables need to be considered, since they can have a considerable effect on the residuals of particular schools, even though they do not contribute significantly to overall variance reduction. It should be noted that having controlled for initial ability and SES at level-1, school average initial ability and SES make independent contributions to variance accounted for. This phenomenon is found in all such studies and is known most commonly as the 'Compositional Effect'.

3.2.1. The 'Compositional' effect

Before proceeding to explore further level-2 variables, the question of the so-called 'Compositional' effect, needs to be briefly discussed. In essence, this question asks whether the school level aggregate of a person level variable (such as SES or initial ability) is significantly related to achievement even after controlling for the variable at the individual level. In HLM, this question can be answered by including the variable at both levels in a model as follows:

level-1 $Y_{ij} = \beta_{0j} + \beta_{1j}*(SES) + r_{ij}$, and

level-2 $\beta_{0j} = \Gamma_{00} + \Gamma_{01}*(SESAGG) + u_{0j}$

$\beta_{1j} = \Gamma_{10}$ (3.3.1.a)

(Note: at level-1, SES is centred around the individual school means)

'The compositional effect is the extent to which the magnitude of the organisational-level relationship, β_b, differs from the person-level effect, β_w.' (B&R:121). If β_b and β_w are equal, then no compositional effect is present. By running the above model against each of the criteria in turn, estimates of the personal- and organisational-effects are

generated, whereby $\beta_w = \Gamma_{10}$; and $\beta_b = \Gamma_{01}$. β_c (the compositional effect) is derived by simple subtraction: $\beta_c = \beta_b - \beta_w$. For this sample of New Zealand schools the relevant data are shown in Table 3.7 for both SES (as in 3.3.1.a.) and initial ability (substituting ZSCOR91 and AGGZ91 for SES and SESAGG in 3.3.1.a.). It is very clear from these data that both the social class composition and the ability mix of the school have a very powerful effect on School Certificate performance, substantially larger than their effects at the individual pupil level. This compositional effect is open to a number of interpretations (B&R:12; Willms 1986):

* the effect may be due to the normative effects associated with particular schools - school 'climate', or 'ethos';
* the aggregated variable may be standing as a proxy for, or acts as an index of, some other important school level variables, such as resources or teacher quality, that have not been included in the model;
* it may be an artifact of a poorly measured variable at the individual level, hence the aggregated variable 'picks-up' the extra effect - an argument favoured by Thomas and Mortimore (1994).

Table 3.7: Compositional effects - disentangling personal and organisational effects.

SES	Maths coeff.	Maths s.e.	Science coeff.	Science s.e.	English coeff.	English s.e.
Γ_{00}	83.833	3.744	74.132	3.918	68.489	3.932
$\Gamma_{01} = \beta_b$	-10.523	1.177	-6.560	1.231	-5.382	1.236
$\Gamma_{10} = \beta_w$	-0.991	0.111	-0.973	0.099	-0.983	0.091
$\beta_c = \Gamma_{01} - \Gamma_{10} =$	9.532		5.587		4.399	
Initial Ability	**Maths coeff.**	**Maths s.e.**	**Science coeff.**	**Science s.e.**	**English coeff.**	**English s.e.**
Γ_{00}	51.394	0.764	53.99	0.592	51.973	0.546
$\Gamma_{01} = \beta_b$	18.215	2.410	14.126	1.869	13.097	1.722
$\Gamma_{10} = \beta_w$	6.631	0.215	6.781	0.186	6.715	0.167
$\beta_c = \Gamma_{01} - \Gamma_{10} =$	11.584		7.345		6.382	

The terminology for the effect varies ('Compositional', 'Contextual', 'School-mix'), but most interpret its existence as a 'peer group' effect (Willms 1986:225)[9], such that the presence in a school of a (large) critical mass of high ability students from high SES families, with a strong positive valuation towards qualifications, and the worth of academic study, has a powerful, independent effect on the examination performance of *all* students who attend such a school. Conversely, of course, the presence of a large mass of students with a negative valuation of schooling has a depressing effect on the scores of *all* students.

> High concentrations of disadvantaged students can adversely affect the school's ability to maintain the social order and can foment peer cultures that act in opposition to the school's academic aims. (Lee, Bryk & Smith 1993:180)

In addition to the normative effect of the peer group, it has also been argued that teachers' classroom practices are influenced by the characteristics of the student group (Ibid). The social and ability mix of the school will also have an impact on the curriculum and the way students are guided into courses. Thus it is important, where strong compositional effects are found, to consider variables at both levels of the model. Failure to do this would result in a distorting under-estimation of the influence of the variable concerned.

3.2.2. *Type B effects*

The addition of the school level student background variables increases the amount of between school variance that can be accounted for by a knowledge of some of the characteristics of the pupil populations that attend each school. The results of adding the school-level contextual variables to the Type A model reported above, are shown in Table 3.8. In the case of all three criteria the amount of between-school variance that can be accounted for is increased further.

---

9 Goldstein refers to the effect as 'contextual' in that 'they allow us to study the effect of the other students on each individual student' (1987:2).

Table 3.8: Regressions with means as outcomes - SES and additional variables.

Mathematics

Fixed Effect	Coefficient	Standard Error	T-ratio	P-value
For INTRCPT1, B0				
INTRCPT2, G00	51.050951	0.643077	79.385	0.000
EUROPC, G01	0.071280	0.095126	0.749	0.297
PACISPC, G02	-0.064990	0.116851	-0.556	0.337
AGGZ91, G03	-4.083178	4.464100	-0.915	0.258
SESAGG, G04	-7.960970	2.038074	-3.906	0.001
For ZSCOR91 slope, B1				
INTRCPT2, G10	6.518212	0.226239	28.811	0.000
For SES slope, B2				
INTRCPT2, G20	0.044353	0.110943	0.400	0.364
For MAORI slope, B3				
INTRCPT2, G30	-1.814374	0.500502	-3.625	0.002

Random Effect		Standard Deviation	Variance Component	df	Chi-square	P-value
INTRCPT1,	U0	3.69795	13.67483	32	304.16646	0.000
level-1,	R	14.19419	201.47500			

Explaining between-school variance (53.04 - 13.67) ÷ 53.04 = 0.7422 = 74 %

Science

Fixed Effect	Coefficient	Standard Error	T-ratio	P-value
For INTRCPT1, B0				
INTRCPT2, G00	53.767825	0.556984	96.534	0.000
MAORIPC, G01	0.074102	0.075908	0.976	0.243
EUROPC, G02	0.138851	0.052689	2.635	0.016
AGGZ91, G03	1.329790	3.900481	0.341	0.372
SESAGG, G04	-1.614319	1.786403	-0.904	0.261
For ZSCOR91 slope, B1				
INTRCPT2, G10	6.707717	0.196196	34.189	0.000
For SES slope, B2				
INTRCPT2, G20	0.074795	0.096212	0.777	0.290
For MAORI slope, B3				
INTRCPT2, G30	-1.548593	0.434186	-3.567	0.002

Random Effect		Standard Deviation	Variance Component	df	Chi-square	P-value
INTRCPT1,	U0	3.20178	10.25139	32	307.20464	0.000
level-1,	R	12.30924	151.51744			

Explaining between-school variance (31.69 - 10.25) ÷ 31.69 = 0.6765 = 68 %

(cont)

BEST COPY AVAILABLE

(Table 3.8 cont.)

English

Fixed Effect	Coefficient	Standard Error	T-ratio	P-value
For INTRCPT1, B0				
INTRCPT2, G00	51.773376	0.504538	102.615	0.000
MAORIPC, G01	0.216089	0.078953	2.737	0.013
PACISPC, G02	0.049957	0.067330	0.742	0.298
SESAGG, G03	1.142017	1.607510	0.710	0.305
AGGZ91, G04	12.427923	3.595641	3.456	0.002
For ZSCORS1 slope, B1				
INTRCPT2, G10	6.646465	0.175851	37.796	0.000
For SES slope, B2				
INTRCPT2, G20	0.035665	0.086235	0.414	0.362
For MAORI slope, B3				
INTRCPT2, G30	-1.176118	0.389163	-3.022	0.007

Random Effect		Standard Deviation	Variance Component	df	Chi-square	P value
INTRCPT1,	U0	2.90423	8.43453	32	325.07902	0.000
level-1,	R	11.03286	121.72391			

Explaining between-school variance (27.06 - 8.43) ÷ 27.06 = 0.6884 = 69 %

The between-school variance is summarised in Table 3.9, which shows that control of initial ability, SES and ethnicity at the individual level accounts for the larger part of the variance, but that school level variables make a significant contibution, even when the variable has been controlled at the individual level (AGGZ91 and MAORIPC for English), or makes no significant unique contribution at the individual level (SESAGG for Maths and EUROPC for Science).

Table 3.9: Analysis of Between-School Variance Accounted For (percent)

Control	Mathematics	Science	English
Level-1*	48	50	55
Level-2	26	18	14
Total	74	68	69

* From Table 3.4

For all three criterion variables, at least two-thirds (about 70%) of the between-school variance can be accounted for by the characteristics of the students in attendance. Having controlled for the effects of these characteristics, the residuals generated by the model are the estimates of Type B effects.

Table 3.10: Type B Effects: Empirical Bayes Estimates - expressed as effect sizes (residuals as proportion of s.d).

School	Mathematics	Science	English
01	-.065	.157	-.034
02	.123	.208	-.049
03	.163	-.076	-.073
04	.049	.091	.199
05	-.012	-.143	-.030
06	-.066	-.057	.111
07	-.146	-.047	-.136
08	.010	-.131	-.017
10	.041	.148	.012
11	.014	.122	.107
12	-.148	-.057	.047
13	-.172	-.139	.038
14	-.106	-.063	.364
15	.132	.038	.232
16	-.040	-.099	-.350
17	-.293	-.218	-.334
18	.373	.177	.001
19	.223	.181	-.105
20	.056	-.113	.015
21	.018	-.046	.232
22	-.102	-.143	-.025
23	.090	.156	.080
24	-.025	.044	-.061
25	-.196	-.385	.014
26	-.214	-.065	.148
27	-.290	-.207	-.264
28	.153	.149	.240
29	-.004	-.252	-.095
30	.064	.290	-.242
31	.072	.107	.193
32	.247	.171	-.042
33	.120	.206	.028
34	-.064	-.073	-.102
35	.224	.131	-.266
36	-.032	-.034	.013
37	.008	.010	.107
38	-.206	-.034	.044

BEST COPY AVAILABLE

Figure 3: Effect Size: English
Control for Student Variables

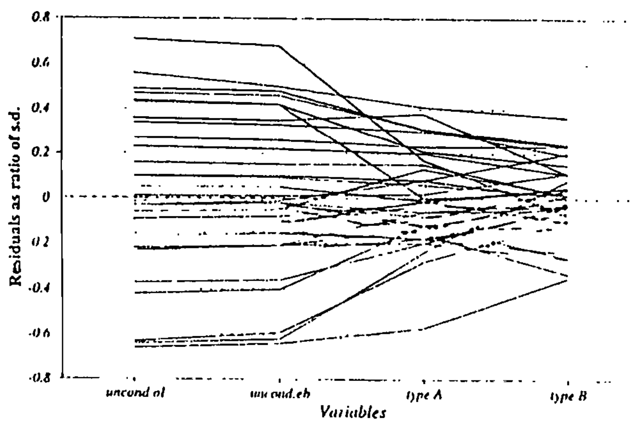


Figure 4:Mathematics Residuals
Type A and Type B Effects (r=0.64)

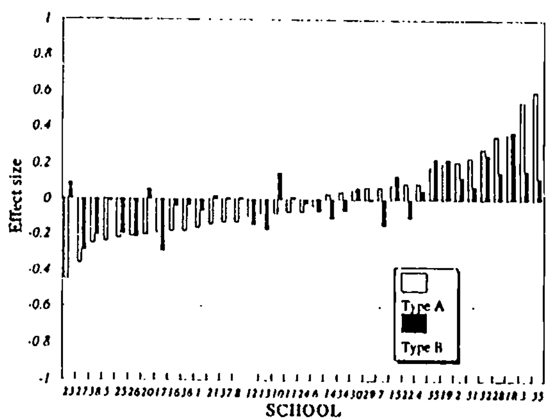


Figure 5:Science Residuals
Type A and Type B Effects (r=0.73)

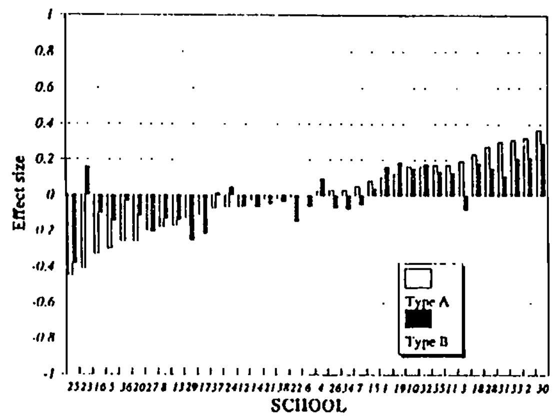
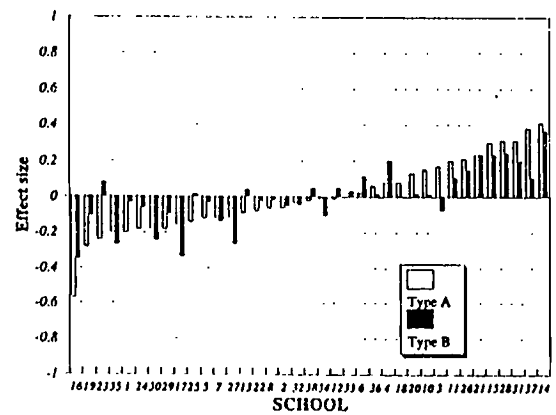


Figure 6: English Residuals
Type A and Type B Effects (r= 0.79)



The Type B residuals are reported in Table 3.10 as 'Effect sizes' - the residual as a proportion of the standard deviation of the criterion variable. By comparing these data with that shown in Table 3.5 it becomes immediately obvious that Type B effects are much smaller than Type A effects, and may take a different direction. School 23 is a good example of a school where almost all the students are from a disadvantaged background for high school achievement, such that the Type A effect is large and negative. However, the Type B effect for the school shows that, given the social-mix of the student population, the school is doing a better than average job, in that more students are getting better marks than students of similar ability and from similar backgrounds in other, similar schools. Thus in terms of the 'value-added' criterion, school 23 is doing as well as, or better, than schools 3 and 33 which have the largest positive Type A effects, and both of which feature prominently in published lists of successful schools (based on unmodified examination pass rates).

It should be noted that the estimate of Type B effects reported here is what Willms (1992:45) would call a 'rough estimate'. The accurate estimation of the effect requires the inclusion of 'direct measures of educational policies and practices' in order to determine their differential effects. Due to some correlation between contextual and school-based factors, it is also possible that some of the effects of policy and practice variables are removed along with the contextual variables. It should also be noted that the Type B residual also contains the effect of errors (sampling, measurement). It is these data (Type B effects) that constitute the basis for the 'value-added' concept in the literature, particularly in Britain (McPherson 1992; Thomas and Mortimore 1994). The effect on the residuals of adjusting the raw mean scores of schools by the initial ability and social and ethnic background of their respective student populations is shown graphically in Figures 1, 2 and 3 for Maths., Science and English results respectively. All scores are shown in relation to the standard deviation of the criterion variable (Effect Size). The Figures show that not only are schools much more similar to each other in terms of their effectiveness than is suggested by raw mean scores, but that those with the highest mean scores are not necessarily the most effective, nor those

with the lowest the least effective. For individual schools, Type A and Type B effects are shown graphically in Figures 4, 5 and 6, which also show the relationship between the two effects (the data in the Figures are ranked by Type A effect).

Figure 1: Effect Size: Mathematics
Control for Student Variables

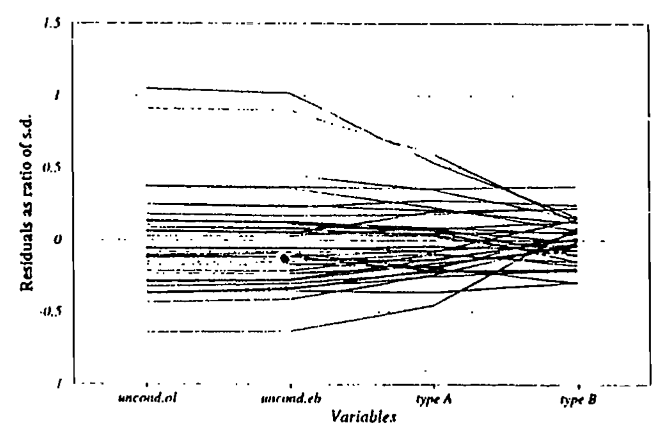
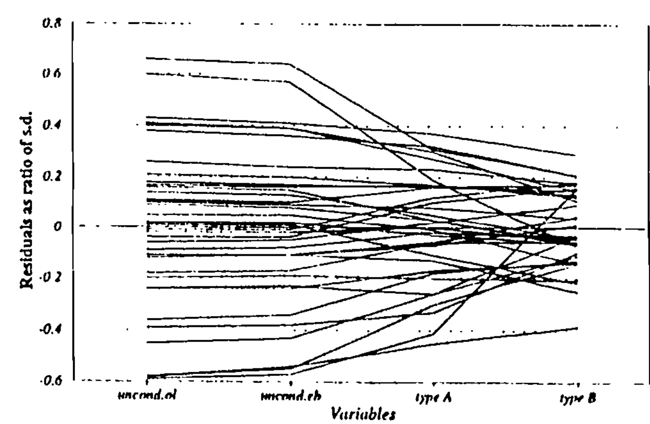


Figure 2: Effect Size: Science
Control for Student Variables



For all three criterion variables a relationship is shown to be present, but it is by no means invariable. School 23 (as already indicated) is an interesting example. Reference to Table 2.1 shows it to be the most disadvantaged school socially, with an exceptionally high proportion of Pacific Island students on its roll. The Type A effect for the school indicates that their pupils do not do well in National examinations of this type. However, the Type B effect shows that for all three criteria the school is doing much better than could be expected from our knowledge of similar students in other schools. Whatever the school is doing (and this school would be a classic site for a qualitative study), it is having a positive effect on the achievements of its students, even though those achievements are well below the national average. In current jargon, it is a 'value-adding' school.

5. Conclusion

In answer to the questions posed in the 'Introduction', the following answers can be given with respect to achievement in the National School Certificate examination:

* 18.2 per cent of total variance in Maths, 14.3 per cent for Science and 14.6 per cent for English, is systematically related to the characteristics of schools as against the inter-individual variability within the population sampled;
* some schools are significantly more effective than others, in that the school means for the School Certificate Examinations in Maths., Science and English vary significantly between schools. This inter-school variability is reduced, but not eliminated, when the initial ability of pupils and the 'school-mix' are taken into account;
* a considerable proportion (around 70%) of the between school variance in all three subject areas is due to the initial ability, social and ethnic characteristics of the pupil population within each school - the Type A effect varies, depending on the initial ability level of the pupils (see Table 3.5);
* further, controlling for these factors at the school level as well (to produce an estimate of Type B effects) produces a dramatic shrinkage in the differences

between schools, and changes in the rank order of effective schools, from the order in which they can be placed on the basis of raw scores only.

Bibliography

Aitkin, M. and Longford, N. (1986) 'Statistical modeling issues in school effectiveness studies' *Journal of the Royal Statistical Society, Series A* 149(1), 1-43.

Bryk, Anthony S. and Raudenbush, Stephen W. (1987). 'Application of hierarchical linear models to assessing change' *Psychological Bulletin* 101(1), 147-158.

Bryk, Anthony S. and Raudenbush, Stephen W. (1992) *Hierarchical Linear Models: Applications and Data Analysis Methods* (Newbury Park (Calif.): Sage Publications Inc.)

Bryk, Anthony S., Raudenbush, Stephen W. and Congdon, Richard T. (1994). *HLM/2L/3L* Computer programmes and User's Reference Guide. (Chicago: Scientific Software International)

Goldstein, H.I. (1986). 'Multilevel mixed linear model analysis using iterative generalized least squares' *Biometrika* 73(1), 43-56.

Goldstein, H.I. (1987). *Multilevel Models in Educational and Social Research* (London: Oxford University Press)

Harker, Richard and Roy Nash (1994) *Supplementary report to 1994 Phase I Report: Preliminary Analysis of School Effects Through the Use of Hierarchical Linear Models.* (Wellington: Ministry of Education)

Kerlinger, Fred N. and Elazar J. Pedhazur (1973) *Multiple Regression in Behavioural Research.* (New York: Holt, Rinehart and Winston).

Kreft, I.G., de Leeuw, J., and Kim, K. (1990). *Comparing Four Different Statistical Packages for Hierarchical Linear Regression: Genmod, HLM, ML2, and VARCL* (Statistics Series No. 50) (Los Angeles: University of California)

Lee, Valerie E., Anthony S. Bryk and Julia B. Smith. (1993) 'The organization of effective secondary schools.' In Linda Darling-Hamilton (Ed.) *Review of Research in Education, 19.* (Washington: American Educational Research Association).

McPherson, A. (1992) *Measuring Added Value in Schools.* (U.K.:National Commission on Education)

Nash, Roy and Harker, Richard. (1994) *Progress at School: Final Report, Phase I* (Wellington: Ministry of Education)

Paterson, L. (1991) 'An introduction to multilevel modelling', in Raudenbush and Willms (eds) *Op.cit.*

Raudenbush, S.W. and J.D. Willms (eds) (1991) *Schools, Classrooms, and Pupils: International Studies of Schooling from a Multilevel Perspective.* (San Diego: Academic Press)

Rutter, M. (1983) 'School effects on pupil progress: research findings and policy implications.' *Child Development* 54(1), 1-29.

Smith, David J. and Sally Tomlinson (1989) *The School Effect: A Study of Multi-Racial Comprehensives.* (London: Policy Studies Institute)

Thomas, Sally and Peter Mortimore (1994) *Report on Value Added Analysis of 1993 GCSE Examination Results in Lancashire.* (London: Curriculum Studies Department, Univ. of London Institute of Education)

Willms, J.Douglas. (1986) 'Social class segregation and its relationship to pupils' examination results in Scotland' *American Sociological Review* 51, 224-241.

Willms, J.Douglas. (1992) *Monitoring School Performance* (London: The Falmer Press)